INTELLECTUAL PROPERTY
INDIA

सत्यमेव जयते

GOVERNMENT OF INDIA
PATENT OFFICE
MINISTRY OF COMMERCE AND INDUSTRY
Department of Industrial Policy and Promotion

It is hereby certified that annexed here to is a true copy of **Provisional Specification and Drawings** of the patent application as filed and detailed below:-

Date Of Application	:	**23/01/2003**
Application No	:	**64/MAS/2003**
Applicants	:	**M/s. Ittiam Systems Private Limited, an Indian Company of 4th & 5th Floors, Consulate-I, Richmond Road, Bangalore – 560 025.**

In witness there of
I have here unto set my hand

Dated this the 29th day of May 2007
8th day of Jyaistha, 1929(Saka)

By Authority of
THE CONTROLLER GENERAL OF PATENTS,
DESIGNS AND TRADE MARKS.

(V. RENGASAMY)
ASSISTANT CONTROLLER OF PATENTS & DESIGNS

PATENT OFFICE
INTELLECTUAL PROPERTY RIGHTS BUILDING
G.S.T. ROAD, GUINDY
CHENNAI – 600 032.

FORM 2

The Patents Act, 1970

[39 of 1970]

Provisional Specification

[See section 10]

1. **"TRANSFORM BASED PERCEPTUAL AUDIO CODERS EMPLOYING IMPROVED QUANTIZATION SCHEME"**

2. **(A) ITTIAM SYSTEMS PRIVATE LIMITED**

   (B) 4th & 5th FLOORS, CONSULATE-I, 1 RICHMOND ROAD, BANGALORE – 560 025

   (C) NATIONALITY – INDIAN

The following specification describes the nature of this invention: -

TRANSFORM BASED PERCEPTUAL AUDIO CODERS EMPLOYING IMPROVED QUANTIZATION SCHEME

Field of the Invention

[0001] The present invention in general relates to quantization techniques used in perceptual audio coders and more specifically to the quantization schemes employed in MPEG 1/2 Layer 3 Audio Coding (MP3) and MPEG 2/4 Advanced Audio Coding (AAC).

Background of the Invention

[0002] Uncompressed CD quality audio requires 1.4 Mbps (Megabits per second) for transmission or storage of stereo music. Advances in audio coding techniques have reduced the bandwidth and storage requirements of high fidelity audio (1.4 Mbits/sec) by a factor of 10-15. These audio coding techniques rely on the principle of human auditory masking to remove the components, which are irrelevant for human perception.

[0003] Perceptual audio coders standardized by ISO MPEG committee have become very popular over the years and are widely employed for audio storage and transmission applications. MPEG is a working group of ISO/IEC in charge of the development of standards for coded representation of digital audio and video. Established in 1988, the group has produced MPEG-1, the standard on which such products as Video CD and MP3 are based, MPEG-2, the standard on which such products as Digital Television set top boxes and DVD are based, MPEG-4, the standard for multimedia for the fixed and mobile web. MPEG committee designs algorithms for the compression (informative section) and specifies the bit stream format exactly (normative section). The informative section is specified more loosely, leaving room for developers to innovate.

[0004] MP3 and AAC algorithms are based on sophisticated psycho-acoustic models and achieve compression by giving less importance to frequencies that are perceptually irrelevant. According to listening tests conducted by expert committees, codecs such as MP3 and AAC provide transparent quality at compression ratios between 10–15.

[0005] The encoding process in perceptual audio coders is compute intensive and requires processors with high computation power to perform real-time encoding. The quantization module of the encoder takes up significant part of the encoding time.

[0006] These audio encoding techniques are commonly used in applications like Digital Audio Broadcasting, ISDN transmission for broadcast contribution and distribution purposes, archival storage for broadcasting, accompanying audio for digital TV (DVB, Video CD, ARIB), Internet streaming, Portable audio, Storage and exchange of music files on computers, Content based Storage and Retrieval, Digital AM Broadcasting, Digital Television, Set-Top Box and DVD, Infotainment, Mobile Multimedia, Real Time Communications ,Streaming Audio-Video on the Internet / Intranet ,Studio and Television Post-production ,Surveillance and Virtual Meeting ,Delivery of audio for wireless distribution - via 3G or Bluetooth and many such applications.

<u>Brief Description of the Drawing Figures</u>

[0007] FIG 1 is a block diagram of Perceptual Audio Encoder.

[0008] FIG 2 is a flow chart of the inner iteration loop of the Quantization Scheme.

[0009] FIG 3 is a flow chart of the outer iteration loop of the Quantization Scheme.

[0010] FIG 4 is a flow chart of the NEW quantization scheme

<u>Summary of the Invention</u>

[0011] Fig 1 illustrates an audio encoding process. In conventional audio coding data is processed frame by frame. A time to frequency transformation, Modified Discrete Cosine Transform (MDCT) 100 is performed to get the spectral lines 103. The psychoacoustic block 101 mimics the human perception system to determine the masking thresholds (distortion thresholds) 102 for groups of neighboring spectral lines referred as one scale factor band. The psycho acoustic block (typically) gives a set of thresholds that indicate the levels of Just Noticeable Distortion (JND), if the quantization noise introduced by the coder is above this level then it is audible. As long as the Signal to (quantization) Noise Ratio (SNR) of the spectral bands are lesser than the Signal to Mask Ratio (SMR) the quantization noise cannot be perceived.

[0012] The spectral lines in these scale factor bands are then non-uniformly quantized 104 and noiselessly coded (Huffman coding) 106 to produce a compressed bit-stream.

[0013] In MPEG Audio encoders (MP3 or AAC) a major portion of the processing time is spent in the quantization module 104 as the process is carried out iteratively. In the conventional method of quantization two loops are run in order to satisfy perceptual and bit rate criteria.

[0014] Prior to quantization the incoming spectral lines are raised to a power of ¾ (Power law Quantizer) 301 so as to provide a more consistent SNR over the range of quantizer values. The Quantizer uses different values of step size for different scale factor bands depending on the distortion thresholds set by the psychoacoustic block.

[0015] The two iterative loops are run over the spectral lines, the loops are referred to as outer loop (distortion measure loop) 300 and inner loop (bit rate control loop) 200. In the inner loop the quantization step-size is increased 205 in order to fit the spectral lines within the given bit-rate. The iterative process involves modifying the step-size (referred to as the global gain, as it is common for the spectrum) till the spectral lines fit in the specified number of bits 204. The outer loop checks for the distortion caused in the spectral lines on a band-by-band basis, 302 and increases quantization precision for bands that have distortion above JND. The precision is raised through step sizes referred to as local gains 306. The process repeats till both bit-rate and distortion conditions meet. The global gain k and the set of local gains r is sent to the decoder along with quantized spectral lines.

Disadvantages of this Quantization Scheme

[0016] The implementation in the quantization scheme involves two iterative loops, each iteration involves quantization, noiseless coding and inverse-quantization to find the best possible match .The code book search mechanisms in noiseless coding and complex mathematical operations in quantization & dequantization stages make this a computationally intensive block. A significant portion of the processing time in encoding is spent in the quantization modules.

New & Improved Quantization Scheme

[0017] This invention details a new quantization scheme, which reduces complexity by eliminating the outer loop 300 completely. This reduces the complexity of the encoder, while maintaining the same quality as a conventional encoding scheme. The quantization step size is broken apart into two parts. The global gain, common to all spectral bands which is used to control the bit rate and local gains which are specific to each band. In order to maintain perceptually minimal quantization noise, the bands which are given more weightage by the psychoacoustic block are quantized with finer step-sizes than the less important ones. In the new scheme the step-size for bands (local gains) are calculated initially and global gain varied to meet the bit-rate criteria.

[0018] The new quantization approach described above reduces the complexity of MPEG Layer 3 and Advanced Audio Coding by 20%-50%. This facilitates real time encoding of audio data at low bit rates on processors/platforms that do not have significant processing power (e.g. mobile multimedia platforms). The invention can be employed in any application involving real time encoding of audio using the dual-loop quantization scheme.

Description of the Invention

[0019] In the quantization scheme used in MP3 and AAC the inner loop [200] (FIG 2) & outer loop [300] (FIG 3) go on iteratively to meet the bit rate and distortion criteria. Under best case conditions the loops will terminate when all bands have distortion below threshold estimated by the psycho-acoustic model [304]. Such conditions typically occur at high bit rates (over 96kbps / channel). Using the above approach at medium/low bit rates will lead to many outer loop iterations before reaching (one of many) exit conditions.

[0020] The problem is more severe at lower bit rates when it's not possible to maintain the quality (SNR below SMR), the loops run many times before ending at some compromised quality these exit conditions are specific to the implementation, one possible effect is that quantization noise is not spread uniformly in all the scalefactor bands i.e. some bands are more severely distorted than others; these numerous iterations add severely to the processing time.

[0021] The present invention performs the quantization process by eliminating the outer loop completely, as compensation for that it does a noise shaping of the spectral lines on a band by band basis by using their local gains 402. The guiding principle is to maintain (as far as possible) the ratio of SMR to SNR constant for all the bands. Two criteria are chosen:

- High sensitivity bands i.e. those having low SMR values should be given more precision as compared to bands having larger SMR values.
- In order to desensitize the bands further to the effects of quantization the local gains of the bands are modified inversely in proportion to their energy content with respect to the frame energy.

The precision in both cases is controlled using the local gains.

[0022] Therefore, the initial step before performing the inner (only) loop is to set the local gains 401 in each band according to the following two criteria: -

1. A low value of SMR will imply a high value of local gain and vice versa.
2. A low value of band energy with respect to the total energy content in the frame will imply a high local gain and vice versa.

The value of the local gain is derived from the energy ratios and SMRs. The equation for setting the local gains which has been arrived at is

$$K_b = -(\text{int})(\alpha * \text{log2}(\text{en}(b) / \text{sum\_en}) + \beta * \text{log2}(\text{SMR}(b)))$$

Where,

K_b is the local gain for band b
log2() is the logarithm to base 2
en(b) is the band energy in band b
sum_en is the total energy in the frame
SMR(b) is the psychoacoustic threshold for band b

$\alpha = 0.6$ and $\beta = 0.4$ are implementation dependant constants (and their values are derived based on experimental results)

α measures weightage due to energy ratio and β the weightage due to SMRs.

[0023] After carrying out a shaping of the spectrum by allotting different amounts of precision to different scale factor bands depending upon band energy ratios and SMRs, the inner loop runs to satisfy bit rate 200. The noise shaping performed is assumed to have taken care of (relatively) meeting the distortion criteria for the bands.

[0024]. It can be clearly noted from the above steps, that the new quantization scheme fully eliminates the outer iteration loop (steps 302 to 309 in FIG 3). This results in significant reduction in the complexity of the Quantizer and hence the encoder. The performance benefit is more pronounced at lower bit rates (< 96kbps), where the distortion loop (outer loop) runs for multiple iterations in a conventional quantization scheme.

[0025] The new quantization scheme while reducing the complexity by anywhere between 20% - 50% maintains the same quality as the conventional quantization scheme. As a measure of quality the MOS (Mean Opinion Score) is measured using the Perceptual Audio Quality Evaluation tool (based on ITU-R BS .1387). The MOS scores for few audio files from SQAM (Sound Quality Assessment material has been provided below.

SQAM Audio Clip	MOS for MP3 Encoder with conventional Quantization Scheme			MOS for MP3 Encoder with new Quantization Scheme		
	64 Kbps	96 Kbps	128 Kbps	64 Kbps	96 Kbps	128 Kbps
frer07_1 / music	4.91	5	5	5	5	5
spme50_1 / Male speech	2.64	4.1	4.7	2.02	3.68	4.58
trpt21_2 / air instrument	2.77	3.65	4.51	2.82	3.58	4.31

Notes:

- All audio clips shown above are stereo files at 44.1khz sample rate
- MOS measured using the EAQUAL tool (a public domain tool based on ITU-R BS.1387 specification)

Performance Benefits of the proposed quantization Scheme

[0026] The table below summarizes the performance improvements (speedup) achieved for the invention. The speedup has been by taking the ratio of the CPU time (measured as cycles) taken for the original MP3 Encoder (with conventional quantization) and CPU time of the encoder with the new quantization scheme

Speedup of the Encoder =
(CPU time with conventional quantization) / (CPU time with new quantization)

Speedup of the Quantizer Module =
(CPU time of conventional quantization module) / (CPU time with new quantization module)

SQAM Audio Clip	Speedup in the MP3 Encoder with New Quantization Scheme			Speedup in the Quantization module with new scheme.		
	64 Kbps	96 Kbps	128 Kbps	64 Kbps	96 Kbps	128 Kbps
frer07_1 / music	2.11	1.49	1.62	4.71	3.57	2.81
spme50_1 / Male speech	3.28	2.56	2.08	7.16	4.92	3.61
trpt21_2 / air instrument	2.89	2.47	1.89	6.04	4.76	3.25

Notes:

- All audio clips shown above are stereo files at 44.1khz sample rate
- CPU time measured as cycles using Quantify tool
- Encoder CPU time measured for stereo processing

Applications

[0027] The new quantization scheme while reducing the complexity by anywhere between 20% - 50% maintains the same quality as the conventional quantization scheme .As this algorithm gives significant reduction in the complexity, it can be used in embedded systems used mobile devices and PDAs. This scheme can be used in applications like Digital audio/video broadcasting, storage, video telephony, audio conferencing and interactive multimedia services.

DATED THIS THE 24th DAY OF JANUARY 2003

Binoy Banerjee
AGENT FOR THE APPLICANT
LEX ORBIS
INTELLECTUAL PROPERTY ATTORNEY
B-1/39, MALVIYA NAGAR
NEW DELHI - 110017

NAME OF THE APPLICANT
ITTIAM SYSTEMS PRIVATE LIMITED

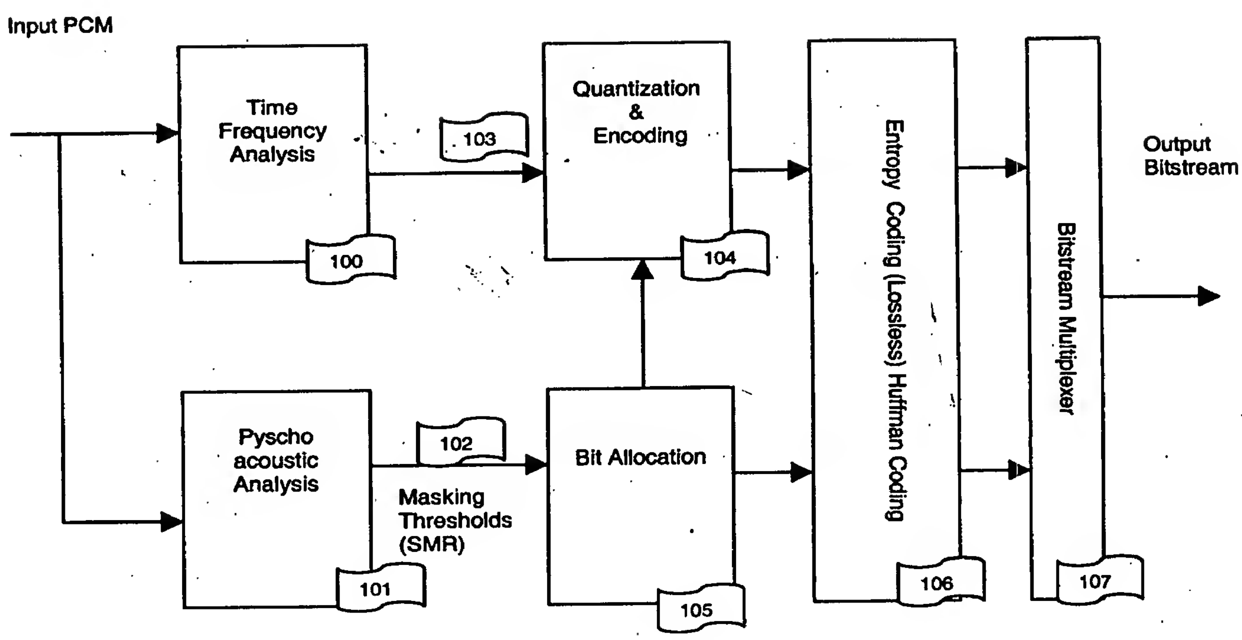

FIG 1

Srinjoy Banerjee
AGENT FOR THE APPLICANT
LEX ORBIS
INTELLECTUAL PROPERTY ATTORNEYS
B-1/39, MALVIYA NAGAR
NEW DELHI - 110 017

NAME OF THE APPLICANT
ITTIAM SYSTEMS PRIVATE LIMITED

TOTAL NO. OF SHEE
SHEET NO. 2 OF

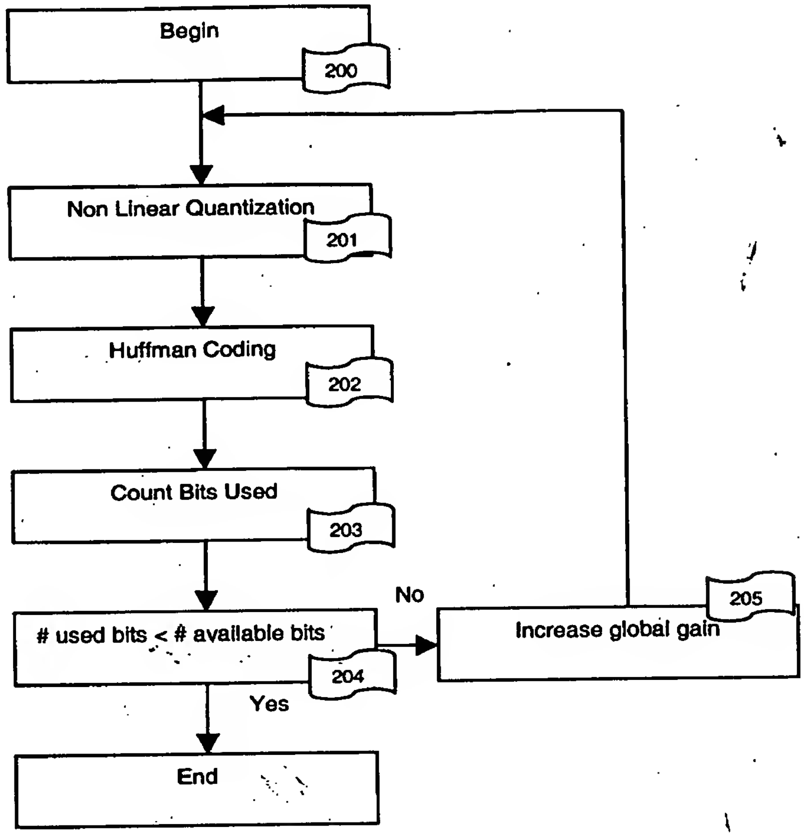

FIG 2

Shinjoy Banerjee
AGENT FOR THE APPLICANT
LEX ORBIS
INTELLECTUAL PROPERTY ATTORNEYS
B-1/39, MALVIYA NAGAR
NEW DELHI - 110017

IAME OF THE APPLICANT
TTIAM SYSTEMS PRIVATE LIMITED

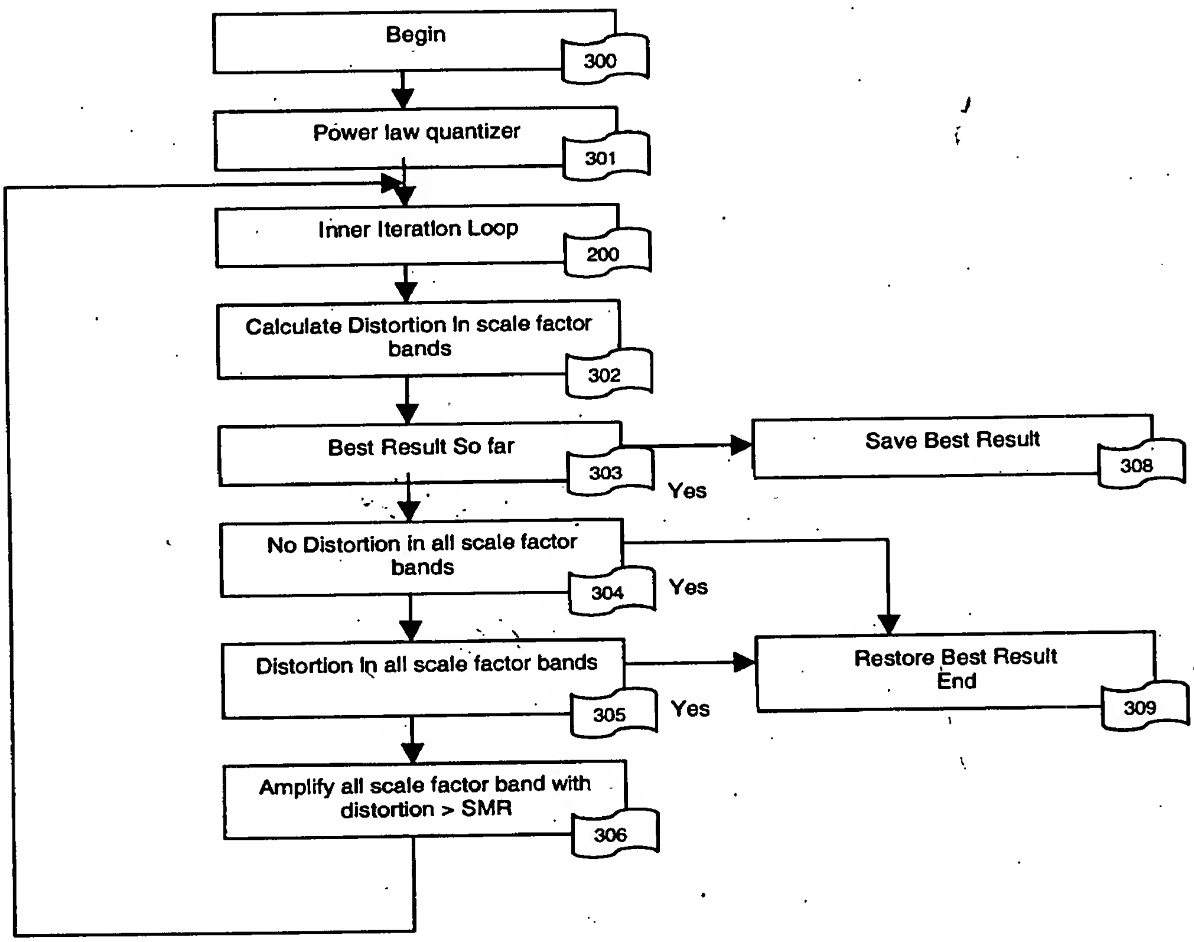

FIG 3

Srinjoy Banerjee
AGENT FOR THE APPLICANT
LEX ORBIS
INTELLECTUAL PROPERTY
ATTORNEYS
B-1/39, MALVIYA NAGAR
NEW DELHI - 110017

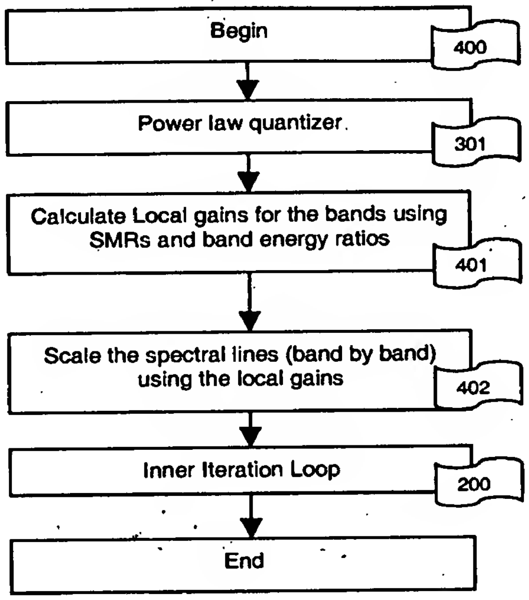

FIG 4

Srinjoy Banerjee
AGENT FOR THE APPLICANT
LEX ORBIS
INTELLECTUAL PROPERTY
ATTORNEYS
B-1/39, MALVIYA NAGAR
NEW DELHI - 110017